Population Genomics Reveals Incipient Speciation, Introgression, and Adaptation in the African Mona Monkey (*Cercopithecus mona*)

Adeola Oluwakemi Ayoola,[1,2,†] Bao-Lin Zhang,[1,*,†] Richard P. Meisel,[3] Lotanna M. Nneji,[1,4] Yong Shao,[1] Olanrewaju B. Morenikeji,[5,6] Adeniyi C. Adeola,[1,4] Said I. Ng'ang'a,[1,2] Babafemi G. Ogunjemite,[7] Agboola O. Okeyoyin,[8] Christian Roos,[9] and Dong-Dong Wu[1,10,11,*]

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

[2]Kunming College of Life Science, University of the Chinese Academy of Sciences, Kunming, Yunnan, China

[3]Department of Biology and Biochemistry, University of Houston, Houston, TX

[4]Sino-Africa Joint Research Center, Chinese Academy of Sciences, Kunming, Yunnan, China

[5]Department of Biomedical Sciences, Rochester Institute of Technology, Rochester, NY

[6]Department of Biology, Hamilton College, Clinton, NY

[7]Department of Ecotourism and Wildlife Management, Federal University of Technology, Akure, Nigeria

[8]National Park Service Headquarters, Federal Capital Territory, Abuja, Nigeria

[9]Gene Bank of Primates and Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

[10]National Resource Center for Non-Human Primates, Kunming Primate Research Center, and National Research Facility for Phenotypic & Genetic Analysis of Model Animals (Primate Facility), Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

[11]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan, China

[†]These authors contributed equally to this work.

***Corresponding authors**: E-mails: zhangbaolin@mail.kiz.ac.cn; wudongdong@mail.kiz.ac.cn.

Associate editor: Rasmus Nielsen

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

Article

Abstract

Guenons (tribe *Cercopithecini*) are the most widely distributed nonhuman primate in the tropical forest belt of Africa and show considerable phenotypic, taxonomic, and ecological diversity. However, genomic information for most species within this group is still lacking. Here, we present a high-quality de novo genome (total 2.90 Gb, contig N50 equal to 22.7 Mb) of the mona monkey (*Cercopithecus mona*), together with genome resequencing data of 13 individuals sampled across Nigeria. Our results showed differentiation between populations from East and West of the Niger River ~84 ka and potential ancient introgression in the East population from other mona group species. The *PTPRK*, *FRAS1*, *BNC2*, and *EDN3* genes related to pigmentation displayed signals of introgression in the East population. Genomic scans suggest that immunity genes such as *AKT3* and *IL13* (possibly involved in simian immunodeficiency virus defense), and *G6PD*, a gene involved in malaria resistance, are under positive natural selection. Our study gives insights into differentiation, natural selection, and introgression in guenons.

***Key words:* guenons, genome sequencing, incipient speciation, introgression, natural selection.**

Introduction

African nonhuman primates are exceptionally diverse, yet knowledge on their genetics, behavior, and ecology is largely limited, which has hindered their conservation and utilization as human disease models. In Africa, guenons (tribe *Cercopithecini*) are one of the most diverse groups of nonhuman primates and show great variation in ecology, behavior, and morphology. Several factors have been illustrated as potential drivers in the evolutionary radiation of guenons (Tosi et al. 2004, 2005). For instance, some guenons are terrestrial, whereas the majorities are arboreal forest-dwellers utilizing different forest canopy levels in sub-Saharan Africa (Gautier et al. 2002). Thus, previous studies (Disotell and Raaum 2002; Guschanski et al. 2013) have stated that speciation in guenons may be associated with forest cover change and specifically forest refugia that occurred in the last 10 Ma.

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

Further, hybridization has also been identified as an important mechanism in the evolution of many lineages, including guenons (Allen et al. 2014; Detwiler 2019; van der Valk et al. 2020), which could have resulted in genetic introgression across species.

Among the five genera of this tribe (i.e., *Cercopithecus*, *Miopithecus*, *Allenopithecus*, *Erythrocebus*, and *Chlorocebus*), the genus *Cercopithecus* is the largest, consisting of 26 species that exhibit variation in genetics, morphology, ecology, behavior, and social organization (Grubb et al. 2003; Anandam et al. 2013; Zinner et al. 2013). Previous studies based on mitochondrial and nuclear sequences (Tosi et al. 2006) have illustrated their phylogeography and identified several species groups, including the *dryas*, *diana*, *mitis*, *cephus*, *hamlyni*, *neglectus*, and *mona* groups (Grubb et al. 2003; Chatterjee et al. 2009; Guschanski et al. 2013). The mona species group has attracted growing attention (Oates 2011; Okekedunu et al. 2014; Onadeko et al. 2014) and is composed of several species or subspecies. Similarly, this group is widely distributed across the African Guineo-Congolian rainforest, with a range extending across Ghana to Cameroon (Oates 2011; Okekedunu et al. 2014). Mona group species live in several types of forest areas within urban, peri-urban, and wild populations (Okekedunu et al. 2014). They usually coexist in sympatry with other primates and often form polyspecific troops, allowing interspecific or even intergeneric hybridization (Gartlan and Struhsaker 1972; Goodwin 2007). Thus, this group provides an ideal model system to investigate the role of genetic introgression in primate evolution. Thus far, no genomic information has been reported for the mona group, which has made it difficult to delineate the evolutionary history and adaptive evolution of this species group.

The advent of large-scale genome sequencing has facilitated investigations into the evolution and ecology of many wild animals, particularly nonhuman primates (Gibbs et al. 2007; Locke et al. 2011; Scally et al. 2012; Prado-Martinez et al. 2013; Carbone et al. 2014; Rogers and Gibbs 2014; Worley et al. 2014; Rogers et al. 2019; Zhang et al. 2019; van der Valk et al. 2020). Herein, we de novo assembled the genome of *Cercopithecus mona* and utilized large-scale genomic data to investigate population differentiation, demographic history, natural selection, and genetic introgression in this species. Interestingly, population genomic analyses revealed a divergence between the East and West Central (WC) populations ~84 ka and identified several immunity genes under positive natural selection. We also found some genes related to pigmentation displaying signals of genetic introgression in the East population from an unknown species. This study should help improve our understanding of the evolution and genetics of the African mona species group.

Results

De Novo Assembly of *C. mona* Genome by Nanopore Sequencing

We sequenced the genome of a female *C. mona* individual using the long-read sequencing platform Oxford Nanopore PromethION. In total, ~156-Gb Nanopore reads and ~151-Gb Illumina Hiseq reads for correcting sequencing errors were generated. The assembled genome was 2.90 Gb with a contig N50 equal to 22.7 Mb (supplementary table S1, Supplementary Material online). The completeness score evaluated by Benchmarking Universal Single-Copy Orthologs was 93.2% (supplementary table S2, Supplementary Material online), indicating that the genome sequence showed high quality and continuity for downstream analyses. Transposable elements (TE) occupied 41.3% of the mona genome, 7.5% of which were long terminal repeats (supplementary table S3, Supplementary Material online). Gene prediction from multiple approaches identified 23,408 protein-coding genes, which is within the range identified in other primate assemblies in the NCBI database. Of these, we identified 20,826 gene families, 8,574 of which were single copies across seven closely related primates (supplementary fig. S1, Supplementary Material online). Estimation of the divergence time based on single-copy gene families suggested that *C. mona* diverged from *Chlorocebus sabaeus* ~9.44 Ma (95% confidence interval [CI] = 7.7–10.5 Ma, fig. 1). This time-frame is similar to that estimated by Pozzi et al. (2014) (~9.03 Ma) but slightly older than other estimates (e.g., Perelman et al. 2011; Guschanski et al. 2013), which may be the result of different data sets and analytical approaches. As comparative genomics analyses with other primate species to understand long-term evolution are outside the scope of this study, we only focused on within species variation to reveal recent evolutionary history of *C. mona* in the following sections.

Population Structure

To investigate the population structure of *C. mona* in Africa, we sequenced the genomes of 13 additional *C. mona* individuals sampled across different coordinates in three geographical regions (East, Central, and West) of Nigeria and one *Cercopithecus mitis* individual as the outgroup (fig. 2A). These individuals were sequenced to an average depth of 24.5-fold (from 22 to 31) (supplementary table S4, Supplementary Material online). After mapping to the reference genome and filtering low-quality variants using the BWA-GATK programs (see Materials and Methods), we obtained 46.6 million high-quality single-nucleotide polymorphisms (SNPs) for downstream analyses. The phylogenetic tree based on all concatenated SNPs suggests that the *C. mona* individuals were composed of two distinct groups: that is, East and WC groups, separated by the Niger River (fig. 2A and B, bootstrap value = 100). The WC lineage could be further subdivided into WCa and WCb lineages without a clear geographic barrier. This topology was supported by trees inferred from coalescent methods (STAR) (Liu et al. 2009) and ASTRAL (Mirarab et al. 2014) based on 10,000 random neutral regions (supplementary fig. S2, Supplementary Material online). Similarly, the principal component analysis (PCA) separated the East and WC groups along the first eigenvector, which explained 12.90% of total genetic variance (fig. 2C), with the second eigenvector identifying WCa and WCb explaining 9.97% of the variance. The optimal number of genetic clusters inferred by STRUCTURE analysis was 3

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

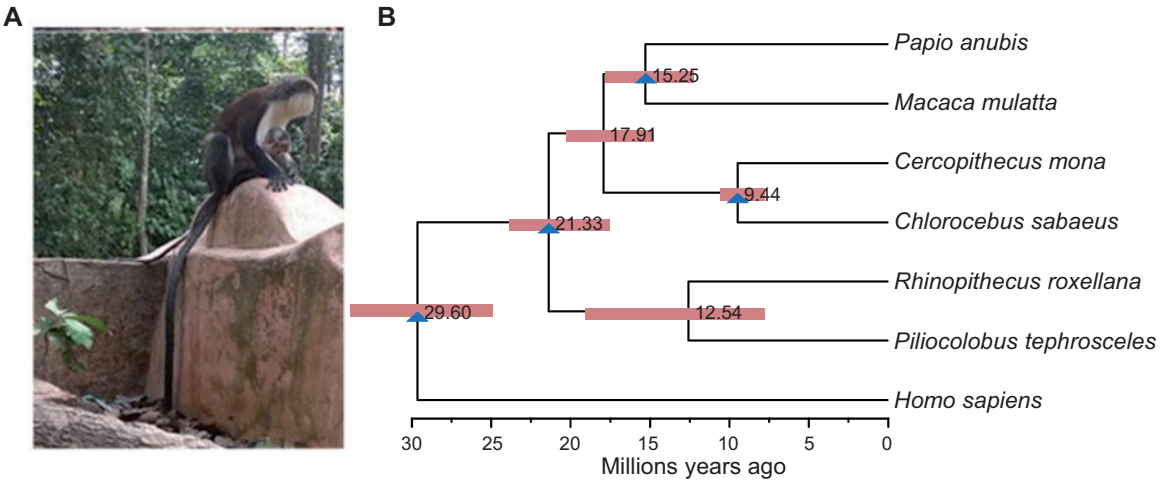

FIG. 1. (*A*) Image of typical *Cercopithecus mona* taken at Osun-Osogbo Sacred Grove protected area. (*B*) Comparative genomics analysis among *C. mona* and six related primates. Node bars refer to 95% confidence intervals of divergence time and blue triangles indicate fossil-based calibration points (see Materials and Methods).

($K = 3$; supplementary fig. S3, Supplementary Material online) using 58,596 unlinked SNPs (fig. 2*D*).

Demographic History and Population Differentiation

We first used the generalized phylogenetic coalescent sampler (G-PhoCS) program (Gronau et al. 2011) to infer divergence times, ancestral effective population sizes, and gene flow rates among the East, WCa, and WCb lineages. Based on comparison of the six models (supplementary fig. S4, Supplementary Material online), Model 3 best fitted the data (Akaike Information Criterion [AIC] = 2,732.30, supplementary tables S5–S7, Supplementary Material online), in which gene flow occurred between WCa and WCb, as well as between the East lineage and the ancestor of the WC population. We further used *D*-statistic analyses to evaluate genetic introgression and found no significant signal of admixture between the East and WCa or WCb samples (Z score $< |3|$, supplementary fig. S5 and table S8, Supplementary Material online). This suggests ancient gene flow between the East population and an ancestor of the WC population but limited recent gene flow between the East and WC populations after divergence between the WCa and WCb lineages. The total migration rate of gene flow from East to WC was 20.3%, which was nearly the same as the opposite direction (20.1%) (fig. 2*E*). The divergence time estimated from the best-fit model showed that the East lineage separated from the WC lineage ~84.4 ka (95% CI = 63.1–104.3 ka), and the divergence between the two WC lineages occurred ~6.7 ka (95% CI = 1.1–14.0 ka) (fig. 2*E* and supplementary table S6, Supplementary Material online).

Pairwise Sequentially Markovian Coalescent (PSMC) analyses showed a similar population size change trajectory for all lineages until ~70 ka (fig. 2*F*). Thereafter, the East lineage experienced a great and prolonged decline in effective population size, whereas the WC lineages expanded 30–40% at first and then experienced a great decline ~30 ka. Demographic history inferred from the Multiple Sequentially Markovian Coalescent (MSMC2) method showed a similar pattern as that of PSMC but exhibited population expansion for all lineages in recent years (<20 ka, supplementary fig. S6, Supplementary Material online). Considering that MSMC2 is more accurate than PSMC when reconstructing recent history, we suspect the population expansions in recent years by MSMC2 may be more reliable. In addition, the separation time estimated by MSMC2 between the East and WC linages was roughly 40–90 ka (supplementary fig. S7, Supplementary Material online), similar to the G-PhoCS results (fig. 2*E*).

Rapid Evolution of Immunity Genes

To identify genes showing rapid evolution, we first calculated genetic differentiation between the East and WC populations based on the population branch statistic (PBS) approach (see Materials and Methods). We focused on windows with extremely high divergence (top 1% PBS value). As a result, we identified 211 regions with a total length of 12.75 Mb, with the longest fragment being 300 kb and the shortest being 50 kb. From these highly divergent regions, 174 protein-coding genes were identified. Gene enrichment analysis indicated that many immunity pathways were significantly enriched ($P < 0.05$, supplementary table S9, Supplementary Material online), for example, Toll-like genes that encode transmembrane proteins, and which recognize unique pathogen-associated molecular structures and play an important role in adaptive immune responses (Descamps et al. 2012). To validate the signal of positive selection on these genes, we performed further analyses to calculate Pi (nucleotide diversity) and cross-population extended haplotype homozygosity (XP-EHH). We identified ten protein-coding genes that showed extremely high PBS and XP-EHH values (top 1%) and extremely low Pi values (low 1%) in the East population compared with the WC population. Among these ten genes, we identified two genes related to immunity against simian immunodeficiency virus (SIV) infection: that is, AKT serine/threonine kinase 3 (*AKT3*) and interleukin 13 (*IL13*) (fig. 3). Moreover, we found that *G6PD* (encoding glucose-6-

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

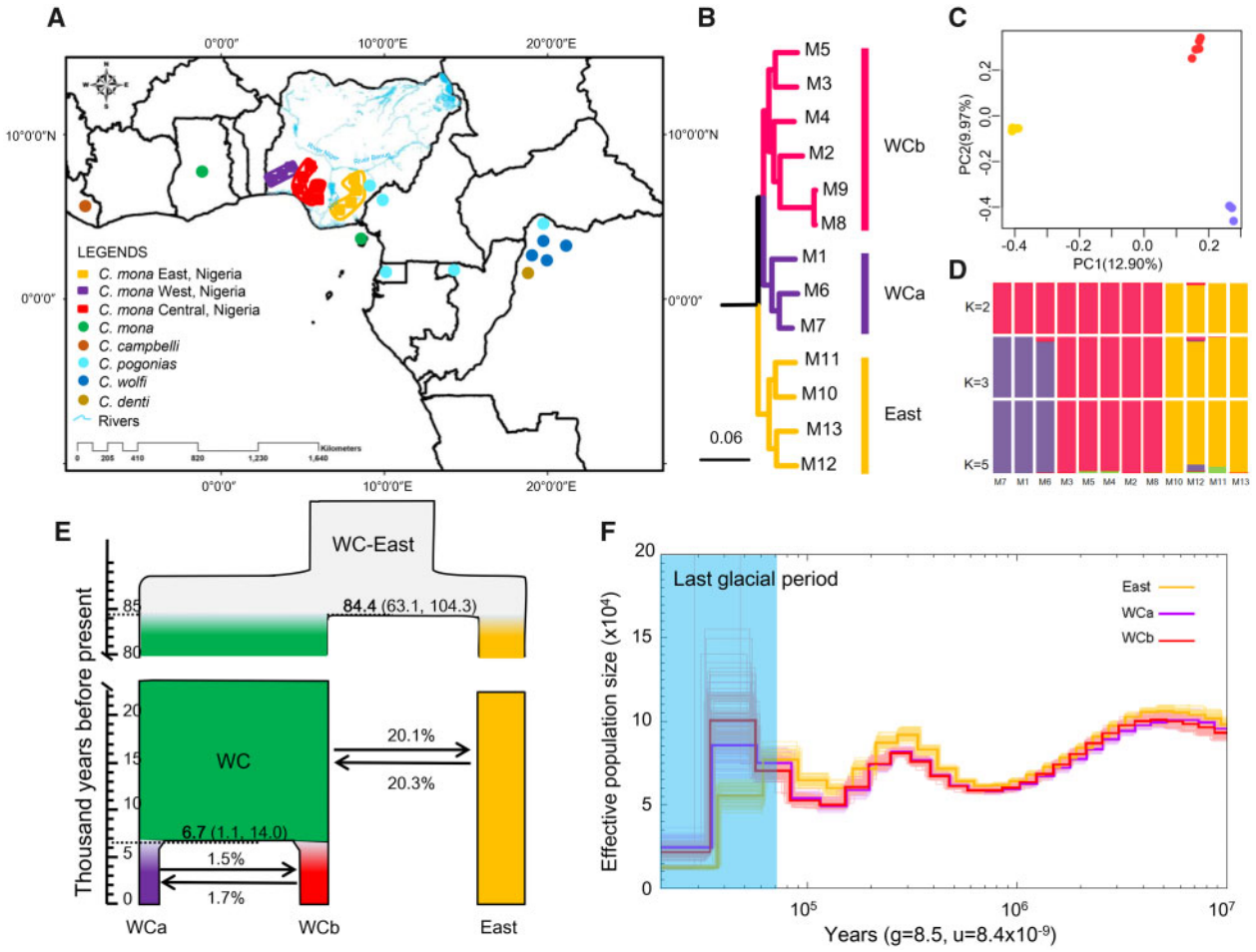

FIG. 2. Population structure and demographic history of *Cercopithecus mona*. (*A*) Sample locations. Rectangles represent sample sites of *C. mona* used for genome resequencing in this study, and their colors correspond to clades recovered from phylogenetic and population structure analyses. Circle dots refer to mitochondrial genome sequences of other species downloaded from NCBI (supplementary table S10, Supplementary Material online). The geographic locations were described in the original study (Guschanski et al. 2013). (*B*) Maximum-likelihood tree of *C. mona* based on all concatenated SNPs. WCa (purple), WCb (red), and East (yellow) indicate three clades referring to West Central clades a and b and East clade, respectively. Outgroup is *Cercopithecus mitis* (not shown). (*C*) PCA results. (*D*) STRUCTURE analyses with $K = 2$–5. (*E*) Demographic history inferred from G-PhoCS results. Widths of branches are proportional to effective population size (Ne). Horizontal dashed lines denote posterior estimates for divergence times in thousand years before present, associated mean values are shown in bold, and 95% confidence intervals are given in parentheses. Arrows and numbers indicate direction and percentage of gene flow. (*F*) PSMC analyses (using a generation time of 8.5 years and an autosomal mutation rate of 8.415×10^{-9} per base pair per generation) show dynamic changes in effective population size. Last Glacial Period (20–70 ka) is shaded in light blue.

phosphate dehydrogenase), a well-known gene related to malaria infection, evolved under positive selection in the East population (fig. 3). Similarly, evolutionary studies suggest that recent selection has operated on *G6PD*-deficient alleles in modern humans, thought to be the result of their protective effect against malaria (Ruwende et al. 1995; Sabeti et al. 2002; Tripathy and Reddy 2007; Louicharoen et al. 2009).

Phylogenetic Analysis Reveals Introgression of Mitochondrial Genomes in East Population from Other Species

Phylogenetic reconstructions based on the mtDNA genome using Bayesian inference and maximum-likelihood (ML) analyses obtained highly supported tree topologies (supplementary fig. S10 and table S11, Supplementary Material online), similar to topologies recovered elsewhere (Guschanski et al. 2013). The topology supported pervasive genetic introgression among African guenons, as previously observed (Guschanski et al. 2013). Particularly, within the *C. mona* species complex, *C. mona* individuals from the East population clustered together with *Cercopithecus pogonias* with very high bootstraps values and Bayesian posterior probabilities (PPs; node 12 in supplementary fig. S10 and table S11, Supplementary Material online). In contrast, all previous studies based on mitochondrial DNA, X-chromosome sequences, Y-chromosome sequences, and karyotypes have indicated a phylogenetic topology of (*C. mona*, (*Cercopithecus wolfi*, *C. pogonias*)) (fig. 4) (Tosi et al. 2005; Moulin et al. 2008; Tosi 2017). The parsimonious explanation for the unusual phylogenetic tree based on the *C. mona* mitochondrial genome is genetic introgression of the East population with other species (likely *C. pogonias*). This is unsurprising given

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

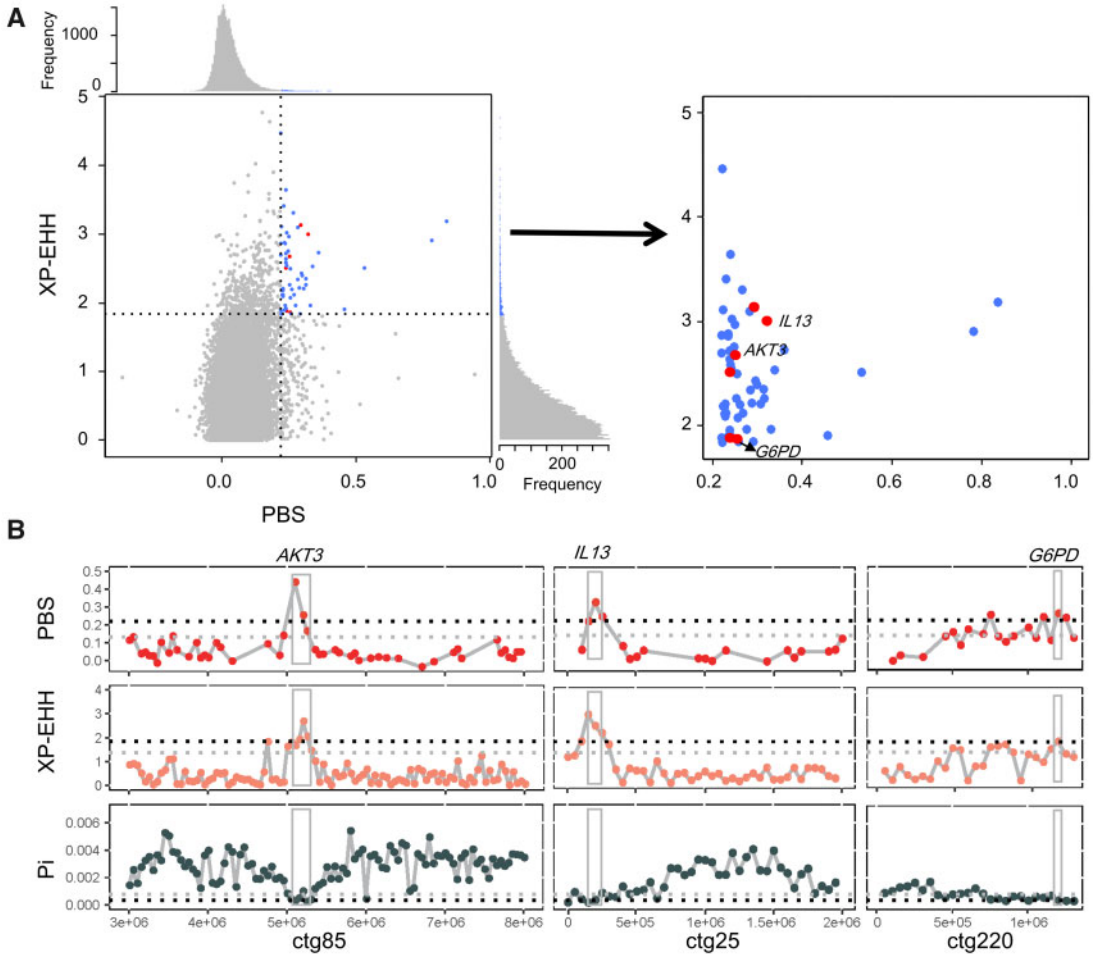

FIG. 3. Selected genes in *Cercopithecus mona* East lineage. (*A*) Scatter plots of XP-EHH against PBS. Data points in blue correspond to genomic windows in top 1% PBS (=0.2190) and top 1% XP-EHH (=1.8369) ratio distribution. Windows also in low 1% Pi ratio distribution are marked in red, which represent candidate selective regions. Genes involved in SIV and malaria resistances are given. (*B*) Genes with strong selective sweep signals in *C. mona* East lineage. Gray and black dashed lines in PBS and XP-EHH indicate top 5% and 1% ratio distributions, respectively, whereas in Pi indicate lower 5% and 1% ratio distributions, respectively.

C. pogonias shows geographical overlap with the East population of *C. mona* (fig. 2*A*). Divergence times between the East and WC populations inferred from the MCMCTREE and BEAST methods were 7.39 Ma (95% highest probability density: 4.00–14.44) and 5.65 Ma (95% highest probability density: 4.72–6.50), respectively (fig. 4 and supplementary figs. S10 and S11, Supplementary Material online), which were much older than that obtained from the nuclear genome (fig. 2*E*). Introgression may also have occurred in WCa or WCb given their older divergence time of mtDNA than nuclear genome. However, our evidence to support this scenario is limited and we cannot exclude other possibility, such as methodology differences, in estimating divergence time.

Identification of Introgression Regions in East Population

Given the observed ancient hybridization of the mitochondrial genomes of the *C. mona* East population, we used the Hidden Markov Model (HMM) to deduce and categorize archaic introgression segments in the genomes of the East individuals (Skov et al. 2018). Using the constant mutation rate of 0.99e-09 per base pair (bp) per year derived from this study, this method inferred divergence between the East and WC populations occurring ~111.7 ka (fig. 5*A*). However, this is a very rough estimate, as described in the original study (Skov et al. 2018), because the method only considers mutations, not gene flow or population size change (Skov et al. 2018). The archaic introgression length decreased when more strict PP cutoff values were applied. After applying a cutoff value of 0.98, the introgression length for different individuals ranged from 8.36 to 8.56 Mb (supplementary fig. S12 and table S12, Supplementary Material online). From these results, we identified an overlap region of ~3.7 Mb, which may harbor candidate introgression genes shared among the East samples.

We randomly selected the same number of genomic regions (3,700 genomic regions with 1 kb in length) and then compared the nucleotide distance (dxy) between random and the putatively archaic introgression regions. Results showed that the latter exhibited a 2.08-fold higher nucleotide divergence than the former (0.0097 vs. 0.0046, supplementary fig. S13, Supplementary Material online). From these

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

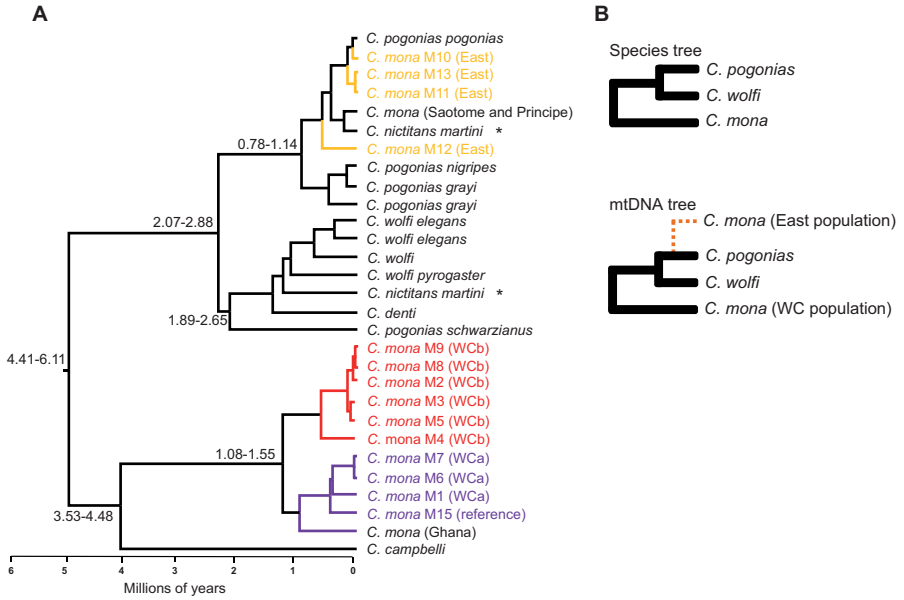

Fig. 4. Phylogenetic relationships reveal hybridization of East population of *Cercopithecus mona* with others. (*A*) Phylogenetic tree of mitochondrial genomes. Details on phylogeny are provided in supplementary figure S10, Supplementary Material online. GenBank ID is provided in supplementary table S10, Supplementary Material online. Numbers beside nodes indicate 95% CIs of divergence time in unit of million years. Tip labels of *C. mona* consist of species codes followed by locality codes: East (Yellow) and West (WCa, Purple; WCb, Red). Individuals in black were downloaded from NCBI (supplementary table S10, Supplementary Material online). The two samples marked with asterisk are likely specimen mix-up as described in Guschanski et al. (2013). (*B*) Species tree according to previous studies based on mitochondrial DNA, X-chromosome sequences, Y-chromosome sequences, and karyotypes (Tosi et al. 2005; Moulin et al. 2008; Tosi 2017). Below is relationship according to mitochondrial phylogeny in (*A*).

overlapping introgression regions, we identified 231 genes (supplementary table S13, Supplementary Material online). We only focused on the longest introgression segments because short archaic segments may be false positives (Skov et al. 2018). From the introgression regions longer than 10 kb, three genes were retrieved, that is, *VWA5B1* (von Willebrand factor a domain containing 5B1), *PTPRK* (protein tyrosine phosphatase receptor type k), and *FRAS1* (Fraser extracellular matrix complex subunit 1). The function of *VWA5B1* is not well documented in the literature, but the latter two genes are reportedly related to pigmentation in humans and other animals (Novellino et al. 2003; Sturm 2009). We constructed a haplotype network for these genes and observed a clear pattern in which the potential introgressed haplotype in the East population showed a large distance to other haplotypes (i.e., *PTPRK* in fig. 5*B*). In addition, among the 231 genes located at the introgressed regions, we also found two other genes, that is, *BNC2* (fig. 5*C*) and *EDN3*, which are related to pigmentation, showing signatures of genetic introgression in the East population (supplementary table S13, Supplementary Material online).

Discussion

The establishment of high-quality reference genomes and population genomic data for primate species is necessary for study on primate evolution. The draft genome of *C. mona* presented here, with high-quality genome assembly and gene annotation, will serve as an ideal reference data point for future genome sequencing and comparative genomic studies. Our research is the first to evaluate the evolutionary history of *C. mona* from close geographical areas using mitochondrial and nuclear genomes based on next-generation sequencing.

Incipient Speciation between WC and East Populations of *C. mona*

We took advantage of the high-depth whole-genome data set to infer the demographic history of *C. mona* using G-PhoCS, as well as changes in the effective population size using PSMC and MSMC2. Population genomics revealed that the WC and East populations diverged ~84 ka, with no recent gene flow between them. The changes in present and ancestral effective population size corroborated long-term differentiation between the WC and East populations. These population genetic features rejected the hypothesis of panmixia between WC and East populations, similar to previous studies on incipient speciation. For example, Zhou et al. (2018) reported incipient speciation between two finless porpoise populations that diverged ~40 ka without gene flow. Rivers are physical barriers associated with the distribution of certain taxa and allopatric diversification (Gascon et al. 2000). African rivers are considered important biogeographic barriers for various mammals (Nicolas et al. 2006, 2010), including within the *Crocidura olivieri* complex (Jacquet et al. 2015) and in other African mammals along the Volta and Congo-Ubangi rivers (Katuala et al. 2008; Nicolas et al. 2010). In Africa, several primate species have also been influenced by rivers, including *Mandrillus* sp., *Cercopithecus erythrotis*, *C. nictitans*, and *C.*

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

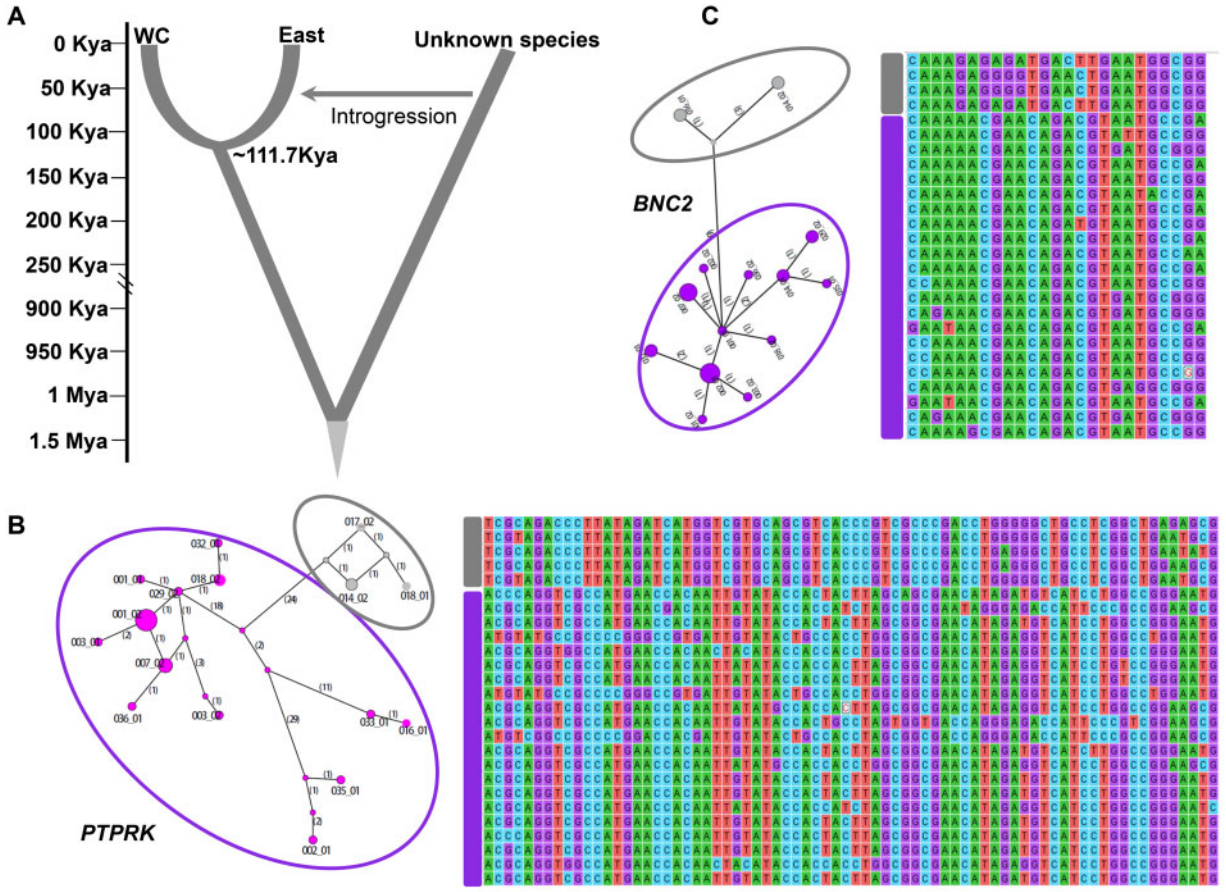

FIG. 5. Analysis of introgressed regions. (*A*) Sketch showing introgression from an unknown *Cercopithecus mona* group lineage into East population of *C. mona*, with scale to the left indicating divergence time ka or Ma. (*B*, *C*) Genomic introgressed regions and median joining haplotype network tree of *PTPRK* and *BNC2*. Number in parentheses on each branch is mutation number. Haplotypes inside gray circle are potential introgressed haplotypes in East population, purple circle embodies other haplotypes. Right of each network is alignment of different haplotypes, with gray rectangle representing potential introgressed haplotypes in East population and purple rectangle representing other haplotypes.

pogonias pogonias (Grubb 2003; Clifford et al. 2004; Anthony et al. 2007; Harcourt and Wood 2012). The Ogooué River is proposed to have delimited the distribution of *Mandrillus sphinx* (Telfer et al. 2003), and the Sanaga River is an important factor influencing patterns of genetic diversity in chimpanzees from Cameroon (Mitchell et al. 2015). We propose that the major reason for this divergence could be because of the Niger and Benue rivers acting as a strong barrier to gene flow (fig. 2A), which have promoted incipient speciation in *C. mona*.

Genetic Introgression in *C. mona* Species Complex

Due to the different patterns of inheritance, simultaneous analysis of nuclear and mitochondrial markers allows for a thorough evaluation of the evolutionary history of *C. mona*. Whole nuclear genomes estimated a divergence time of ~84 ka between the East and WC populations. In contrast, the phylogenetic tree constructed by mitochondrial genome sequences revealed a higher differentiation between the East and WC populations, with a divergence time of ~6.05 Ma. The East population clustered with *C. pogonias*, whereas the WC population clustered with other species (fig. 4). Unusual topology of the mtDNA sequences has been documented in other primate species (Tosi et al. 2004; Kuang et al. 2019), and in a recent study on Dryas monkeys (*Cercopithecus dryas*) (van der Valk et al. 2020). The parsimonious explanation is the probable occurrence of ancient hybridization. Prior studies have revealed frequent genetic introgression events between closely related primates (Detwiler et al. 2005; Tung and Barreiro 2017; Malukiewicz 2019), and confirmed the presence of admixture signatures in previously described primate hybrid zones (Cortes-Ortiz et al. 2007; Tung et al. 2008; Charpentier et al. 2012) and in new primate species complexes (Gligor et al. 2009). For example, de Manuel et al. (2016) identified introgression from bonobos to eastern and central chimpanzee subspecies, which are the closest geographically to the range of modern bonobos. Similarly, yellow baboons sampled with Anubis baboons near their range contain significantly higher levels of introgressed Anubis ancestry than those farther away (Charpentier et al. 2012; Wall et al. 2016).

In light of the species topology, the parsimonious explanation for the unusual phylogenetic tree based on mitochondrial genomes is the occurrence of genetic introgression in the East population with other species (fig. 4), although we cannot entirely exclude other scenarios. Thus, we further asked

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

whether any functional genes or pathways have been introgressed in East population individuals from unknown species. Our evolutionary analyses identified several introgressed genes likely related to pigmentation, including *PTPRK* and *FRAS1*. *PTPRK* is a protein tyrosine phosphatase-κ gene showing expression in human primary keratinocytes, which are the source of peptides recognized by T cells in patients with melanoma (Novellino et al. 2003). The *FRAS1* gene encodes extracellular matrix proteins, which play critical roles in regulating membrane adhesion in the epidermal basement during embryonic development (Short et al. 2007). In addition to these two genes, another two genes (*BNC2* and *EDN3*) involved in pigmentation also exhibited signals of genetic introgression. *BNC2* encodes a protein that functions in skin color saturation and is associated with pigmentation in East Asian populations (Hider et al. 2013). *BNC2* expression and mutation levels are associated with facial pigment spots and have been widely studied in mice (Smyth et al. 2006; Vanhoutteghem et al. 2009) and zebrafish (Lang et al. 2009; Patterson and Parichy 2013). *EDN3* is involved in melanocyte development and pigmentation (Saldana-Caboverde and Kos 2010), and a variant of this gene can cause dermal hyperpigmentation in chickens (Dorshorst et al. 2011).

Haplotype networks of these pigmentation genes supported high divergence of introgression haplotypes in the East population with others from the WC population (fig. 5*B* and *C*). However, the functional consequences of the introgression of these pigmentation genes remain unclear. We propose that introgression of genes related to pigmentation may be related to color evolution in *C. mona*. Further ecological field surveys are necessary to detect any phenotypic differences between the East and WC populations, which will help in understanding the consequences of genetic introgression.

Rapid Evolution of *G6PD* in African Guenons

In this study, we identified the *G6PD* gene as evolving under positive selection in the *C. mona* populations. *G6PD* provides instructions for the production of a glucose-6-phosphate dehydrogenase enzyme and its deficiency is related to malaria resistance. In modern humans, *G6PD*-deficiency mutations are widespread in specific geographic areas and are associated with resistance to malaria infection by *Plasmodium* parasites (Ruwende et al. 1995; Howes et al. 2012). Evolutionary studies have identified selection on the *G6PD*-deficient allele, believed to be the result of its malaria-protective effect in modern humans, like African and Southeast Asian (Ruwende et al. 1995; Sabeti et al. 2002; Tripathy and Reddy 2007; Louicharoen et al. 2009). Malaria is common in most African regions and one of the leading causes of death globally (Greenwood and Mutabingwa 2002; WHO 2015). It is also challenging for many animals in Africa. Until now, the biology of malaria infection in African guenons has been largely unstudied. For example, whether some populations are infected by malaria parasites or are resistant to malaria is unclear. We propose that rapid evolution of *G6PD* may be related to malaria resistance in some special African guenon populations; however, this requires further experimental validation.

Rapid Evolution of SIV Infection-Related Genes in African Guenons

As a diverse group of lentiviruses, SIVs infect many primate species in sub-Saharan Africa (Pandrea et al. 2008; Ansari et al. 2014). Studies have suggested that primates have been infected with SIVs for millions of years (Keele et al. 2009; Rudicell et al. 2010; Etienne et al. 2011). Guenon species, such as *C. mona*, *C. nictitans*, and *C. cephus*, can be infected with SIVmon, SIVmus, and SIVgsn, respectively (Dazza et al. 2005; Schmokel et al. 2011). Studies suggest that these species-specific SIVs are the result of simultaneous host diversification, with concurrent splits occurring in SIV and primate lineages (host-dependent evolution) (Sharp et al. 2001). Evidence of this pattern has also been reported in African vervet monkeys (*Chlorocebus tantalus*, *Chlorocebus pygerythrus*, *Chlorocebus sabaeus*, and *Chlorocebus aethiops*) infected with species-specific SIVs. Using genome-wide population genomics, we identified two candidate SIV immunity genes, that is, *AKT3* and *IL13*, which showed evidence of positive natural selection in *C. mona*. The *Akt3* pathway plays a substantial role in the expression of Nef-mediated CCL5 (HIV-1 Nef is a protein expressed early in infection, which plays a major role in downregulation and viral pathogenesis) (Liu et al. 2014) and in different neurological mechanisms, such as proper growth and myelination during the course of infection (Easton et al. 2005; Tsiperson et al. 2013). *IL13* significantly downregulates CD8 expression in HIV-specific CD8(+) T cells. To enhance this downregulation, previous investigations proffered a novel IL-13R cytokine trap vaccine approach for HIV-1 (Ranasinghe et al. 2013). Interestingly, *IL13* has also been reported to play an important role in malaria resistance (Dewasurendra et al. 2012; Maiga et al. 2013). The rapid evolution of *AKT3* and *IL13* may be attributable to an "arms race" between SIV and monkey hosts. Within African guenons, *C. mona* may also be a potential biomedical animal model for future investigations on SIV infection, in addition to vervet monkeys.

Materials and Methods

Taxon Sampling

Animal collection approval and ethical clearance (NPH/GEN/121/XXV/561) for sample collection and transportation were obtained from the National Park Service of Nigeria. Import permission was approved by the National Forestry and Grassland Administration of China. Sampling followed the protocols of animal use authorized by the Kunming Institute of Zoology Animal Care and Ethics Committee. A total of 14 *C. mona* samples from different localities within three geographical regions in Nigeria as well as one *C. mitis* sample were collected for sequencing (fig. 2*A* and supplementary table S4, Supplementary Material online). Preliminary species identification was based on external morphology following previous study (Oates 2011). Samples were obtained via confiscation from poachers in the communities surrounding Nigerian national parks. Tissue samples collected from carcasses were preserved in 95% ethanol at −80 °C.

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

Genome Sequencing and Assembly

One female *C. mona* sample was chosen for reference genome sequencing and assembly. Total genomic DNA was extracted from preserved muscle in ethanol using an QIAGEN extraction kit (Cat#13323, Blood & Cell Culture DNA Mini Kit). After assessing DNA quality, a library was prepared and sequenced on the Oxford Nanopore PromethION long-read sequencing platform. This generated a total of 156.3-Gb genomic data (~54× coverage) with an average read length of 20.5 kb after removing the adaptor sequences and low-quality reads. Before genome assembly, NextDenovo (https://github.com/Nextomics/NextDenovo, last accessed June 2019) was applied to correct the high error rate of Nanopore long reads using the Illumina short reads. The corrected reads were then assembled using wtdbg v1.2.8 (https://github.com/fantasticair/wtdbg-1.2.8, last accessed June 2019), with the options wtdbg-1.2.8 -k 0 -p 23 -S 2, wtdbg-cns -c 3 -k 15, kbm-1.2.8 -k 0 -p 21 –S 2, and wtdbg-cns -k 13 -c 3. Preliminary genome assembly was further calibrated using NextPolish in NextDenovo. We used Benchmarking Universal Single-Copy Orthologs (Simao et al. 2015) to assess the completeness of our genome assembly and to search the assembly for annotated genes conserved among all mammals.

Repeat Elements and Protein-Coding Gene Prediction

Repeat elements in the *C. mona* genome were annotated by both ab initio and homology-based approaches. We first developed a de novo repeat library using RepeatModeler v1.0.8 (Smit et al. 2008) with default parameters then used RepeatMasker v4.0.6 (Benson 1999) to identify known and de novo repeats by searching against the RepBase21.11 database (Bao et al. 2015) and de novo repeat library, respectively. RepeatProteinMask implemented in RepeatMasker was used to identify the TE-relevant proteins. In addition, we employed TRF v4.07 to predict the tandem repeats in the genome using default parameters. All identified repeat elements were classified into different categories (DNA, LINE, long terminal repeat, SINE, Unknown, and Other) according to the repeat database classification. Repeat annotations for the genome were combined into a nonredundant repeat annotation. In total, we predicted 41.32% of bases as TEs in the genome.

Protein-coding genes were also predicted by a combination of homology- and ab initio-based strategies. For homology-based prediction, the protein-coding sequences were obtained from the NCBI database for humans (*Homo sapiens*, GCA_000001405.28), chimpanzees (*Pan troglodytes*, GCA_002880755.3), gorillas (*Gorilla gorilla*, GCA_900006655.3), orangutans (*Pongo abelii*, GCA_002880775.3), and mice (*Mus musculus*, GCA_000001635.8). These protein sequences were then mapped to the *C. mona* genome using TBlastN v2.2.26 (Altschul et al. 1990) with an *E*-value cutoff of 1e-5. Proteins with multiple adjacent hits were connected to each other using genBlastA v1.0.4 (She et al. 2009). The aligned sequences and query proteins were then filtered and transferred to GeneWise v2.4.1 (Birney et al. 2004) to identify spliced alignments. Augustus v3.0.3 (Keller et al. 2011) was used for ab initio prediction, with optimized parameters trained from 1,000 randomly selected homologous genes. Finally, we integrated all gene sets to form a comprehensive and nonredundant gene set using in-house Perl scripts.

Divergence Time Estimation

Gene sequences from six well-assembled primates (fig. 1) were downloaded from the NCBI database for comparative genomic analysis. For genes with numerous transcript isoforms, we only kept the longest for downstream analyses. Treefam (Ruan et al. 2008) was used to estimate gene families. We identified 8,574 one-to-one ortholog genes from these steps. Sequence alignments for each one-to-one ortholog were performed using MAFFT (Katoh et al. 2005). Phylogenetic relationships were constructed using RAxML v.8.1.15 (Stamatakis 2014) under the GTR + GAMMA model. Divergence times were inferred with MCMCTREE in PAML v4.8 (Yang 2007). To calibrate time, four calibrations were used: 1) *Homo sapiens–Macaca mulatta*: 23.0–33.9 Ma (Benton and Donoghue 2007), 2) Colobinae–Cercopithecinae: 8.5–23.03 Ma (Springer et al. 2012), 3) *Macaca–Papionini*: 5.5–23.03 Ma, and 4) a second calibration of *Chlorocebus–Cercopithecus*: 7.60–10.43 Ma (Pozzi et al. 2014). Calibrations were scaled to units of 100 My. The constraints for minimum and maximum bounds were soft, with a default 2.5% probability that bounds could be violated. Markov chain Monte Carlo (MCMC) was performed, with 10,000 generations as burn-in, then sampling every 100,000 generations until a total of 10,000 samples were generated. Two independent runs were executed, and convergence for each run to ensure an effective sample size (ESS) of more than 200 was determined in Tracer v1.6 (Rambault et al. 2014).

Mutation Rate Estimation

The putative one-to-one orthologs between *C. mona* and humans identified above were used for mutation rate estimation. The yn00 program in the PAML 4.9e package (Yang 2007) was used to calculate synonymous distances (*Ks*) between orthologs. The formula "$r = Ks/2t$" (Wang et al. 2020) was applied to estimate the neutral substitution rate, where "t" is the mean divergence time between *C. mona* and humans. After the above steps, we obtained an average mutation rate of 0.99e-09 per bp per year, slightly smaller than the estimates reported by Pfeifer (2017) (0.99e-09 vs. 1.1e-09).

Genome Resequencing and SNP Calling

From the samples (supplementary table S4, Supplementary Material online), 13 *C. mona* individuals and one *C. mitis* individual were used for whole-genome resequencing. Sequence libraries were designed in accordance with the Illumina library preparation framework and then sequenced on the Illumina Hiseq Xten platform to generate 150-bp paired-end reads using a 350-bp library. Raw sequence reads were mapped with BWA-MEM v0.7.12-r1039 (Li 2013) to the reference genome with default parameters. SAMtools v1.9 (Bigham et al. 2009) was used to sort and remove polymerase chain reaction duplicates. To reduce the number of miscalls

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

around insertions/deletions (indels), the Genome Analysis Toolkit (GATK) v.2.6-5 (DePristo et al. 2011) was used to realign the raw gapped alignment. With the command "samtools mpileup -q 20 -Q 20 -C 50 -uDEf," raw SNPs were called using SAMtools from locally realigned BAM files. To acquire high-quality SNPs for downstream analyses, the following filter criteria were used: 1) removal of sites around 6 bp from predicted indels, 2) removal of sites with consensus quality <40, 3) removal of sites with triallelic alleles and indels, and 4) retention of sites present in at least 95% of individuals. The filtered SNPs were phased using a read aware phasing approach implemented in SHAPEIT v2.0 (Delaneau et al. 2013). We used the following parameters to run SHAPEIT in read-conscious phasing mode: 200 conditional states, 10 iterations burn-in, 10 iterations pruning, 50 main iterations, and 0.5-Mb window size.

Population Structure, Gene Flow, and Phylogeny

To exclude closely related samples from population structure analyses, we first estimated pairwise Identity-By-State scores among all samples by using PLINK v 1.90b5 (https://www.cog-genomics.org/plink/, last accessed October 2019). We found that M8 and M9 showed high pairwise genomic similarity (0.4988) thus leaving one of them (M8) in the downstream analyses. We used the Bayesian clustering method in Structure v2.3.4 (Pritchard et al. 2000) to infer the genetic ancestry of each sample by specifying the hypothetical ancestral clusters (K) from 1 to 5. We kept one SNP for each interval of 50 kb to avoid the linkage disequilibrium effect, resulting in 58,596 sites. Five independent runs were conducted, with the first 10,000 MCMC iterations treated as burn-in and the following 50,000 iterations then used. The most probable K value was estimated using Structure Harvester (http://taylor0.biology.ucla.edu/structureHarvester/, last accessed October 2019). PCA was analyzed using the GCTA v.1.24.2 package (Yang et al. 2011). To infer admixture events, we applied the D-statistic method in COMP-D (Mussmann et al. 2019). Based on a consensus-rooted four-taxon phylogeny, this method tests whether there is significantly different allele sharing between a source lineage (P3), and either of two receiving lineages (P1, P2) with reference to the outgroup (O). Here, to examine gene flow between the East and WC populations, we defined samples from the East population as P3, samples from WC as P1 and P2, and *C. mitis* as O.

Phylogenetic relationships were first estimated using the concatenated whole-genome SNPs with RAxML v.8.1.15 under the GTR + GAMMA model, with *C. mitis* as the outgroup (Stamatakis 2014). To minimize the effects of regions with strong natural selection and ancestral incomplete lineage sorting, we performed coalescent-based phylogenetic analyses using STAR (Liu et al. 2009) and ASTRAL v.4.10.8 (Mirarab et al. 2014) based on 10,000 putatively neutral 50-kb genomic windows by filtering out positions with repeat sequences, exons, and the 10-kb regions flanking them on each side. These windows should be at least 100 kb distant from each other. Individual gene trees for each window were constructed using RAxML v.8.1.15. Support values for each node were obtained using 100 rapid bootstrap replicates based on the GTR + GAMMA model.

Demographic History Inference

The full-likelihood approach implemented in G-PhoCS v1.2.2 (Gronau et al. 2011) was used to infer the entire demographic history of *C. mona*, including ancestral population sizes, population divergence times, and migration rates. We randomly collected 1,000 loci (1,000-bp long) from the putatively neutral genomic region identified above for inference. We limited our data set to three individuals within each population due to computational constraints. To evaluate different migration scenarios among the different populations, as well as split time and goodness-of-fit model, six demographic models were established and compared (supplementary fig. S5 and tables S7 and S8, Supplementary Material online). The loci of the ghost population in Model 6 were set as "N." The average log-likelihood value was obtained from MCMC and then used to calculate the AIC $= -2$ log-likelihood + 2P, where P is the number of parameters in the model. The model with the lowest AIC value was chosen as the best model. Every Markov chain was run for 3,000,000 iterations, with the first 1,000,000 iterations removed as burn-in, and thereafter sampling each 50th iteration. For the best model, we randomly choose an additional four independent data sets and reran G-PhoCS to check whether the estimated demographic parameters were stable. Convergence of each run was determined using Tracer v1.6.

Because G-PhoCS assumes a constant population size along phylogenetic branches, which may prevent the capture of gradual population size changes, for example, population bottlenecks, we adopted two other model-flexible methods: that is, PSMC (Li and Durbin 2011) and MSMC2 (Schiffels and Durbin 2014). PSMC and MSMC2 can be complementarily used as the former method exhibits better performance when inferring ancient histories (e.g., more than 20 ka), whereas the latter method is more reliable and well supported for recent population histories (Schiffels and Durbin 2014). The parameters for the PSMC analysis were -N25 -t15 -r5 -b -p "$4 + 25 \times 2 + 4 + 6$." A bootstrapping approach was carried out with 100 replicates to evaluate variation in the inferred effective population size (Ne) trajectories. For MSMC2 analysis, we selected three individuals with the highest sequencing coverage from each group of interest. As MSMC2 requires haplotypes as input, the phased VCF file from SHAPEIT was used. Other input files were prepared using the scripts provided in GitHub (https://github.com/stschiff/msmc-tools, last accessed October 2019) and MSMC2 was run using default parameters. We used a generation time of 8.5 years (Anandam et al. 2013) and a neutral mutation frequency of 0.99e-09 per site per year (as estimated above). A 50% elative cross coalescent rate was used as a rough estimate of divergence time (Malaspinas et al. 2016; Wang et al. 2020).

Mitochondrial Genome Assembling and Alignment

Mitochondrial genome sequence of each individual was obtained from Illumina Hiseq Xten reads by using NOVOplasty 2.4 (Dierckxsens et al. 2017). K-mer was set to

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

33 and one mitogenome of *C. mona* (JQ256979) download from GenBank was used as bait reference. Assembled contigs were annotated with MITOS (http://mitos2.bioinf.uni-leipzig.de/index.py, last accessed October 2019). The reliability of mitochondrial contig assemblies was further validated via BLAST searches against the reference mitgenome (JQ256979). Sequences were aligned using MUSCLE v3.8.31 (Edgar 2004), as implemented in MEGA v7.0.14 (Kumar et al. 2016). Protein-coding genes were translated to amino acid sequences to ensure an open reading frame and avoid amplification of NUMTs (nuclear mitochondrial DNA segments), as NUMTs have many stop codons in the sequences.

Mitochondrial Phylogeny and Divergence Time

Including the outgroup *C. mitis*, we generated 15 near complete mitochondrial genome sequences in this study. For phylogenetic reconstructions, additional sequences were added from GenBank (supplementary table S10, Supplementary Material online) to expand our data sets covering mtDNA genomes from a total of 79 African guenon samples. We did not include the d-loop sequences due to their low quality. We generated alignments for individual loci with MEGA 7.0 (Kumar et al. 2016). Indels and badly aligned positions were removed using standard settings with Gblocks v0.91b (Castresana 2000). The aligned genomes were partitioned into protein-coding genes, noncoding fragments, rRNAs, and tRNAs. The protein-coding genes were further partitioned into first, second, and third codon positions using Split Codons (Stothard 2000). We used PartitionFinder v2.1.6 to evaluate the best partitioning scheme and to determine replacement models for each partition using corrected AIC (AICc) (Lanfear et al. 2012). ML analysis was conducted in RAxML v.8.1.15 with 1,000 bootstrap replications under the best partition scheme (https://embnet.vital-it.ch/raxml-bb/) (Stamatakis 2014). Bayesian inference analysis was performed using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003). We conducted two independent runs using four MCMC runs, each for 10,000,000 generations, with tree sampling every 1,000 generations. Tracer v1.6 (Rambault et al. 2014) was used to assess convergence visually and to verify that the ESS was >200. The posterior distribution of trees was summarized after checking for congruence by removing the first 25% of generations as burn-in. Lineages with Bayesian PP and posterior branch support values higher than 0.95 were considered strongly supported (Erixon et al. 2003).

Divergence time was estimated by MCMCTREE and BEAST v1.8.0 (Drummond and Rambaut 2007), respectively. In the MCMCTREE analysis, we included our data as 17 partitions according the PartitionFinder results. Parameters and calibrations were the same as used in comparative analysis. For BEAST analysis, a lognormal relaxed clock was used to estimate branch lengths, and the Yule model was used for the tree as priors. We performed four independent MCMC runs, each with 20,000,000 million generations, and with trees and parameters sampled every 2,000 generations. The burn-in and convergence of chains were assessed using Tracer v1.6 (Rambault et al. 2014), with a burn-in of 25%, to verify that the ESS was >200. We used LogCombiner v1.8.3 to combine results of the log file independent runs and Tree Annotator 2.4.4 was implemented to create a standard tree with maximum clade credibility using median node heights in BEAST. The tree was visualized, summarized, and edited using FigTree v1.4.2 (Rambaut 2012).

Analysis of Introgression

We adopted a newly developed HMM method (Skov et al. 2018) to infer the signature of archaic introgression in the East *C. mona* group. This approach can identify a high density of foreign nucleotide variants without a reference genome and shows a lower false positive rate than that of other methods, such as S* (Vernot and Akey 2014; Vernot et al. 2016) and Sprime (Browning et al. 2018). The WCa + WCb individuals were treated as an outgroup. The callable regions were identified using the repeat masked genome and then classified into bins of 1,000 bp. The number of variants found in the outgroup was calculated using VCFtools v0.1.11 (Danecek et al. 2011). In 100-kb windows, the background mutation rate was estimated using the variant density of all variants in the outgroup population. HMM was configured with a set of starting parameters based on human parameters (Skov et al. 2018). To ensure accuracy of the final parameter data, we designed the model across five separate runs, varying the starting parameters. The posterior decoding then specified whether consecutive 1-kb windows changed or maintained their condition ("archaic" or "nonarchaic"), dependent on PP.

Inference of High-Divergence Regions and Selection Signals between East and WC Populations

Under topological transformation outlined previously (Wang et al. 2018), genomic differentiation was determined based on PBS (Yi et al. 2010). The PBS value calculates the extent of sequence shift in a population branch as the divergence between it and other branches of a population tree. First, using Weir and Cockerham's method in VCFtools v0.1.11 (Danecek et al. 2011), F_{ST} pairwise statistics were calculated with nonoverlapping 50-kb genomic windows. Negative F_{ST} values were treated as 0. Log-transformed F_{ST} values were used for PBS calculation (Yi et al. 2010; Wang et al. 2018). The upper 1% of each PBS distribution was described as the high divergence regions (HDRs). Typically, regions that have experienced selection will show significantly decreased levels of nucleotide diversity and long-range homozygosity of the haplotype (Sabeti et al. 2006). To check for signatures of selective sweeps in HDRs, we first calculate the nucleotide diversity (Pi) of each window using VCFtools v 0.1.11. The XP-EHH values (haplotype-based statistics) were estimated using selscan v1.1.0a (Szpiech and Hernandez 2014) assuming 1 Mb = 1 cm across the *C. mona* genome. By subtracting the genome-wide mean XP-EHH and dividing by the standard deviation, the XP-EHH values were normalized. The average XP-EHH values of SNPs in each window were calculated by an in-house Perl script. Gene Ontology enrichment was analyzed with DAVID (Database for Annotation Visualization and Integrated Discovery) (Dennis et al. 2003). Only Gene Ontology terms with a P value of <0.05 were considered.

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We are grateful to the Nigerian National Park Service and the Department of Ecotourism and Wildlife management FUTA for the support during the fieldwork. This work was supported by the National Natural Science Foundation of China (31822048), Sino-Africa Joint Research Center, Chinese Academy of Sciences (SAJC201611), and the Animal Branch of the Germplasm Bank of Wild Species of Chinese Academy of Science (the Large Research Infrastructure Funding). D.-D.W. was supported by the Youth Innovation Promotion Association, Chinese Academy of Sciences.

Data Availability

The Nanopore sequence data and Illumina sequencing reads were deposited in the NCBI database under the project accession number PRJNA595456. The *Cercopithecus mona* genome assembly and annotation files were deposited in GWH under accession number GWHALOC00000000 and NCBI under JABBCR000000000. Mitochondrial genome data generated in this study are available in GenBank under accession numbers MT544450–MT544457 and MT936554–MT936560.

References

Allen WL, Stevens M, Higham JP. 2014. Character displacement of *Cercopithecini* primate visual signals. *Nat Commun*. 5(1):4266.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215(3):403–410.

Anandam MV, Bennett EL, Davenport TRB, Davies NJ, Detwiler KM, Engelhardt A, Eudey AAG, Gadsby EL, Groves CP, Healy A, et al. 2013. Species accounts of Cercopithecidae. In: Mittermeier R, Rylands A, Wilson D, editors. Handbook of the mammals of the world. Vol. 3. Primates. Barcelona (Spain): Lynx. p. 628–753.

Ansari A, Silvestri G, Peeters M, Ma D, Liegeois F, Apetrei C. 2014. Simian immunodeficiency virus infections in the wild. In: Ansari A, Silvestri G, editors. Natural hosts of SIV. New York: Elsevier. p. 37–67.

Anthony NM, Johnson-Bawe M, Jeffery K, Clifford SL, Abernethy KA, Tutin CE, Lahm SA, White LJT, Utley JF, Wickings EJ, et al. 2007. The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proc Natl Acad Sci U S A*. 104(51):20432–20436.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6:11.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 27(2):573–580.

Benton MJ, Donoghue PC. 2007. Paleontological evidence to date the tree of life. *Mol Biol Evol*. 24(1):26–53.

Bigham AW, Mao X, Mei R, Brutsaert T, Wilson MJ, Julian CG, Parra EJ, Akey JM, Moore LG, Shriver MD. 2009. Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. *Hum Genomics*. 4(2):79–90.

Birney E, Clamp M, Durbin R. 2004. GeneWise and genomewise. *Genome Res*. 14(5):988–995.

Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM. 2018. Analysis of human sequence data reveals two pulses of archaic Denisovan admixture. *Cell* 173(1):53–61.

Carbone L, Alan Harris R, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J, Meyer TJ, Herrero J, Roos C, Aken B, et al. 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature* 513(7517):195–201.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 17(4):540–552.

Charpentier MJ, Fontaine MC, Cherel E, Renoult JP, Jenkins T, Benoit L, Barthes N, Alberts SC, Tung J. 2012. Genetic structure in a dynamic baboon hybrid zone corroborates behavioural observations in a hybrid population. *Mol Ecol*. 21(3):715–731.

Chatterjee HJ, Ho SY, Barnes I, Groves C. 2009. Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evol Biol*. 9(1):259.

Clifford SL, Anthony NM, Bawe-Johnson M, Abernethy KA, Tutin CE, White LJ, Bermejo M, Goldsmith ML, McFarland K, Jeffery KJ, et al. 2004. Mitochondrial DNA phylogeography of western lowland gorillas (*Gorilla gorilla gorilla*). *Mol Ecol*. 13(6):1551–1565.

Cortes-Ortiz L, Duda TF, Canales-Espinosa D, Garcia-Orduna F, Rodriguez-Luna E, Bermingham E. 2007. Hybridization in large-bodied New World primates. *Genetics* 176(4):2421–2425.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.

Dazza MC, Ekwalanga M, Nende M, Shamamba KB, Bitshi P, Paraskevis D, Saragosti S. 2005. Characterization of a novel vpu-harboring simian immunodeficiency virus from a Dent's Mona monkey (*Cercopithecus mona denti*). *J Virol*. 79(13):8560–8571.

de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, Hernandez-Rodriguez J, Dupanloup I, Lao O, Hallast P, et al. 2016. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* 354(6311):477–481.

Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 10(1):5–6.

Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol*. 4(5):P3.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 43(5):491–498.

Descamps D, Le Gars M, Balloy V, Barbier D, Maschalidi S, Tohme M, Chignard M, Ramphal R, Manoury B, Sallenave JM. 2012. Toll-like receptor 5 (*TLR5*), IL-1beta secretion, and asparagine endopeptidase are critical factors for alveolar macrophage phagocytosis and bacterial killing. *Proc Natl Acad Sci U S A*. 109(5):1619–1624.

Detwiler KM. 2019. Mitochondrial DNA analyses of *Cercopithecus* monkeys reveal a localized hybrid origin for *C. mitis doggetti* in Gombe National Park, Tanzania. *Int J Primatol*. 40(1):28–52.

Detwiler KM, Burrell AS, Jolly CJ. 2005. Conservation implications of hybridization in African cercopithecine monkeys. *Int J Primatol*. 26(3):661–684.

Dewasurendra RL, Suriyaphol P, Fernando SD, Carter R, Rockett K, Corran P, Kwiatkowski D, Karunaweera ND, Consortium M; in collaboration with the 7 MalariaGEN Consortium. 2012. Genetic polymorphisms associated with anti-malarial antibody levels in a low and unstable malaria transmission area in southern Sri Lanka. *Malar J*. 11(1):281.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 45(4):e18.

Disotell TR, Raaum RL. 2002. Molecular timescale and gene tree incongruence in the guenons. In: Glenn ME, Cords M, editors. The guenons: diversity and adaptation in African monkeys. New York: Kluwer Academic/Plenum. p. 27–36.

Dorshorst B, Molin A-M, Rubin C-J, Johansson AM, Strömstedt L, Pham M-H, Chen C-F, Hallböök F, Ashwell C, Andersson L. 2011. A complex genomic rearrangement involving the endothelin 3 locus causes dermal hyperpigmentation in the chicken. *PLoS Genet* . 7(12):e1002412.

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 7(1):214.
Easton RM, Cho H, Roovers K, Shineman DW, Mizrahi M, Forman MS, Lee VM, Szabolcs M, de Jong R, Oltersdorf T, et al. 2005. Role for *Akt3*/protein kinase Bgamma in attainment of normal brain size. *Mol Cell Biol*. 25(5):1869–1878.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5):1792–1797.
Erixon P, Svennblad B, Britton T, Oxelman B. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst Biol*. 52(5):665–673.
Etienne L, Nerrienet E, LeBreton M, Bibila GT, Foupouapouognigni Y, Rousset D, Nana A, Djoko CF, Tamoufe U, Aghokeng AF, et al. 2011. Characterization of a new simian immunodeficiency virus strain in a naturally infected *Pan troglodytes troglodytes* chimpanzee with AIDS related symptoms. *Retrovirology* 8(1):4.
Gartlan JS, Struhsaker TT. 1972. Polyspecific associations and niche separation of rain-forest anthropoids in Cameroon, West Africa. *J Zool*. 168(2):221–266.
Gascon C, Malcolm JR, Patton JL, da Silva MNF, Bogart JP, Lougheed SC, Peres CA, Neckel S, Boag PT. 2000. Riverine barriers and the geographic distribution of Amazonian species. *Proc Natl Acad Sci U S A*. 97(25):13672–13677.
Gautier JP, Drubbel VR, Deleporte P. 2002. Phylogeny of the Cercopithecus lhoesti group revisited: combining multiple character sets. In: Glenn ME, Cords M, editors. The guenons: diversity and adaptation in African monkeys. New York: Kluwer Academic/Plenum. p. 37–48.
Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316(5822):222–234.
Gligor M, Ganzhorn JU, Rakotondravony D, Ramilijaona OR, Razafimahatratra E, Zischler H, Hapke A. 2009. Hybridization between mouse lemurs in an ecological transition zone in southern Madagascar. *Mol Ecol*. 18(3):520–533.
Goodwin RM. 2007. Behaviour and ecology of the Mona monkey (Cercopithecus mona Shreber, 1774) in the Seasonally Dry Lama Forest, Republic of Benin, West Africa. New York: City University of New York.
Greenwood B, Mutabingwa T. 2002. Malaria in 2002. *Nature* 415(6872):670–672.
Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet*. 43(10):1031–1034.
Grubb P, Butynski TM, Oates JF, Bearder SK, Disotell TR, Groves CP, Struhsaker TT. 2003. Assessment of the diversity of African primates. *Int J Primatol*. 24(6):1301–1357.
Guschanski K, Krause J, Sawyer S, Valente LM, Bailey S, Finstermeier K, Sabin R, Gilissen E, Sonet G, Nagy ZT, et al. 2013. Next-generation museomics disentangles one of the largest primate radiations. *Syst Biol*. 62(4):539–554.
Harcourt AH, Wood MA. 2012. Rivers as barriers to primate distributions in Africa. *Int J Primatol*. 33(1):168–183.
Hider JL, Gittelman RM, Shah T, Edwards M, Rosenbloom A, Akey JM, Parra EJ. 2013. Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. *BMC Evol Biol*. 13(1):150.
Howes RE, Piel FB, Patil AP, Nyangiri OA, Gething PW, Dewi M, Hogg MM, Battle KE, Padilla CD, Baird JK, et al. 2012. *G6PD* deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map. *PLoS Med*. 9(11):e1001339.
Jacquet F, Denys C, Verheyen E, Bryja J, Hutterer R, Kerbis Peterhans JC, Stanley WT, Goodman SM, Couloux A, Colyn M, et al. 2015. Phylogeography and evolutionary history of the *Crocidura olivieri* complex (Mammalia, Soricomorpha): from a forest origin to broad ecological expansion across Africa. *BMC Evol Biol*. 15:71.
Katoh K, Kuma K-I, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 33(2):511–518.
Katuala PGB, Kennis J, Nicolas V, Wendelen W, Hulselmans J, Verheyen E, Van Houtte N, Dierckx T, Dudu AM, Leirs H, et al. 2008. The presence of *Praomys*, *Lophuromys*, and *Deomys* species (Muridae, Mammalia) in the forest blocks separated by the Congo River and its tributaries (Kisangani region, Democratic Republic of Congo). *Mammalia* 72(3):223–228.
Keele BF, Jones JH, Terio KA, Estes JD, Rudicell RS, Wilson ML, Li Y, Learn GH, Beasley TM, Schumacher-Stankey J, et al. 2009. Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature* 460(7254):515–519.
Keller O, Kollmar M, Stanke M, Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* 27(6):757–763.
Kuang W-M, Ming C, Li H-P, Wu H, Frantz L, Roos C, Zhang Y-P, Zhang C-L, Jia T, Yang J-Y, et al. 2019. The origin and population history of the endangered golden snub-nosed monkey (*Rhinopithecus roxellana*). *Mol Biol Evol*. 36(3):487–499.
Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 33(7):1870–1874.
Lanfear R, Calcott B, Ho SY, Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol*. 29(6):1695–1701.
Lang MR, Patterson LB, Gordon TN, Johnson SL, Parichy DM. 2009. Basonuclin-2 requirements for zebrafish adult pigment pattern development and female fertility. *PLoS Genet*. 5(11):e1000744.
Li H.2013. Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. eprint arXiv:1303.3997.
Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496.
Liu L, Yu L, Pearl DK, Edwards SV. 2009. Estimating species phylogenies using coalescence times among sequences. *Syst Biol*. 58(5):468–477.
Liu X, Shah A, Gangwani MR, Silverstein PS, Fu M, Kumar A. 2014. HIV-1 Nef induces CCL5 production in astrocytes through p38-MAPK and PI3K/Akt pathway and utilizes NF-kB, CEBP and AP-1 transcription factors. *Sci Rep*. 4:4450.
Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang S-P, Wang Z, Chinwalla AT, Minx P, et al. 2011. Comparative and demographic analysis of orangutan genomes. *Nature* 469(7331):529–533.
Louicharoen C, Patin E, Paul R, Nuchprayoon I, Witoonpanich B, Peerapittayamongkol C, Casademont I, Sura T, Laird NM, Singhasivanon P, et al. 2009. Positively selected *G6PD*-Mahidol mutation reduces *Plasmodium vivax* density in Southeast Asians. *Science* 326(5959):1546–1549.
Maiga B, Dolo A, Toure O, Dara V, Tapily A, Campino S, Sepulveda N, Risley P, Silva N, Corran P, et al. 2013. Human candidate polymorphisms in sympatric ethnic groups differing in malaria susceptibility in Mali. *PLoS One* 8(10):e75675.
Malaspinas AS, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, Bergstrom A, Athanasiadis G, Cheng JY, Crawford JE, et al. 2016. A genomic history of Aboriginal Australia. *Nature* 538(7624):207–214.
Malukiewicz J. 2019. A review of experimental, natural, and anthropogenic hybridization in *Callithrix* marmosets. *Int J Primatol*. 40(1):72–98.
Mirarab S, Bayzid MS, Boussau B, Warnow T. 2014. Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science* 346(6215):1250463.
Mitchell MW, Locatelli S, Sesink Clee PR, Thomassen HA, Gonder MK. 2015. Environmental variation and rivers govern the structure of chimpanzee genetic diversity in a biodiversity hotspot. *BMC Evol Biol*. 15(1):1.
Moulin S, Gerbault-Seureau M, Dutrillaux B, Richard FA. 2008. Phylogenomics of African guenons. *Chromosome Res*. 16(5):783–799.
Mussmann SM, Doglas MR, Bangs MR, Douglas ME. 2019. Comp-D: a program for comprehensive computation of *D*-statistics and

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

population summaries of reticulated evolution. *Conserv Genet Resour*. 16:1–5.
Nicolas V, Akpatou B, Wendelen WIM, Kerbis Peterhans J, Olayemi A, Decher JAN, Missoup A-D, Denys C, Barriere P, Cruaud C, et al. 2010. Molecular and morphometric variation in two sibling species of the genus *Praomys* (Rodentia: Muridae): implications for biogeography. *Zool J Linn Soc*. 160(2):397–419.
Nicolas V, Querouil S, Verheyen E, Verheyen W, Mboumba JF, Dillen M, Colyn M. 2006. Mitochondrial phylogeny of African wood mice, genus *Hylomyscus* (Rodentia, Muridae): implications for their taxonomy and biogeography. *Mol Phylogenet Evol*. 38(3):779–793.
Novellino L, Renkvist N, Rini F, Mazzocchi A, Rivoltini L, Greco A, Deho P, Squarcina P, Robbins PF, Parmiani G, et al. 2003. Identification of a mutated receptor-like protein tyrosine phosphatase kappa as a novel, class II HLA-restricted melanoma antigen. *J Immunol*. 170(12):6363–6370.
Oates JF. 2011. Primates of West Africa. A field guide and natural history. Bogota (Columbia): Panamericana Formase Impress, S. A.
Okekedunu JO, Ogunjemite BG, Adeyemo AI, Olaniyi OE. 2014. Daily activity budget of Mona monkey (*Cercopithecus mona*) in Ibodi Monkey Forest, Osun State. *FUTA J Res Sci*. 2014(2):218–227.
Onadeko A, Olaleru F, Bada O. 2014. The distribution of Mona monkeys (*Cercopithecus mona*, Schreber, 1774) in the University of Lagos Campus, Akoka, Lagos. *Ethiop J Biol Sci*. 13:49–56.
Pandrea I, Sodora DL, Silvestri G, Apetrei C. 2008. Into the wild: simian immunodeficiency virus (SIV) infection in natural hosts. *Trends Immunol*. 29(9):419–428.
Patterson LB, Parichy DM. 2013. Interactions with iridophores and the tissue environment required for patterning melanophores and xanthophores during zebrafish adult pigment stripe formation. *PLoS Genet*. 9(5):e1003561.
Perelman P, Johnson WE, Roos C, Seuanez HN, Horvath JE, Moreira MA, Kessing B, Pontius J, Roelke M, Rumpler Y, et al. 2011. A molecular phylogeny of living primates. *PLoS Genet*. 7(3):e1001342.
Pfeifer SP. 2017. Direct estimate of the spontaneous germ line mutation rate in African green monkeys. *Evolution* 71(12):2858–2870.
Pozzi L, Hodgson JA, Burrell AS, Sterner KN, Raaum RL, Disotell TR. 2014. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Mol Phylogenet Evol*. 75:165–183.
Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. *Nature* 499(7459):471–475.
Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959.
Rambaut A. 2012. FigTree. Available from: http://www.tree.bio.ed.ac.uk/software/figtree. Accessed October 2019.
Rambault A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6 Computer Program. Available from: http://tree.bio.ed.ac.uk/software/tracer/. Accessed October 2019.
Ranasinghe C, Trivedi S, Stambas J, Jackson RJ. 2013. Unique IL-13Ralpha2-based HIV-1 vaccine strategy to enhance mucosal immunity, CD8(+) T-cell avidity and protective immunity. *Mucosal Immunol*. 6(6):1068–1080.
Rogers J, Gibbs RA. 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat Rev Genet*. 15(5):347–359.
Rogers J, Raveendran M, Harris RA, Mailund T, Leppälä K, Athanasiadis G, Schierup MH, Cheng J, Munch K, Walker JA, et al. 2019. The comparative genomics and complex population history of Papio baboons. *Sci Adv*. 5(1):eaau6947.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572–1574.
Ruan J, Li H, Chen Z, Coghlan A, Coin LJ, Guo Y, Heriche JK, Hu Y, Kristiansen K, Li R, et al. 2008. TreeFam: 2008 Update. *Nucleic Acids Res*. 36(Database):D735–D740.
Rudicell RS, Holland Jones J, Wroblewski EE, Learn GH, Li Y, Robertson JD, Greengrass E, Grossmann F, Kamenya S, Pintea L, et al. 2010. Impact of simian immunodeficiency virus infection on chimpanzee population dynamics. *PLoS Pathog*. 6(9):e1001116.
Ruwende C, Khoo SC, Snow RW, Yates SN, Kwiatkowski D, Gupta S, Warn P, Allsopp CE, Gilbert SC, Peschu N, et al. 1995. Natural selection of hemi- and heterozygotes for *G6PD* deficiency in Africa by resistance to severe malaria. *Nature* 376(6537):246–249.
Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419(6909):832–837.
Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. *Science* 312(5780):1614–1620.
Saldana-Caboverde A, Kos L. 2010. Roles of endothelin signaling in melanocyte development and melanoma. *Pigment Cell Melanoma Res*. 23(2):160–170.
Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* 483(7388):169–175.
Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 46(8):919–925.
Schmokel J, Sauter D, Schindler M, Leendertz FH, Bailes E, Dazza MC, Saragosti S, Bibollet-Ruche F, Peeters M, Hahn BH, et al. 2011. The presence of a vpu gene and the lack of Nef-mediated downmodulation of T cell receptor-CD3 are not always linked in primate lentiviruses. *J Virol*. 85(2):742–752.
Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH. 2001. The origins of acquired immune deficiency syndrome viruses: where and when? *Philos Trans R Soc Lond B* 356(1410):867–876.
She R, Chu JS, Wang K, Pei J, Chen N. 2009. genBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res*. 19(1):143–149.
Short K, Wiradjaja F, Smyth I. 2007. Let's stick together: the role of the *Fras1* and *Frem* proteins in epidermal adhesion. *IUBMB Life* 59(7):427–435.
Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
Skov L, Hui R, Shchur V, Hobolth A, Scally A, Schierup MH, Durbin R. 2018. Detecting archaic introgression using an unadmixed outgroup. *PLoS Genet*. 14(9):e1007641.
Smit A, Hubley R, Green P. 2008. RepeatModeler Open-1.0.8. Available from: http://www.repeatmasker.org/RepeatModeler.html. Accessed June 2019.
Smyth IM, Wilming L, Lee AW, Taylor MS, Gautier P, Barlow K, Wallis J, Martin S, Glithero R, Phillimore B, et al. 2006. Genomic anatomy of the *Tyrp1* (brown) deletion complex. *Proc Natl Acad Sci U S A*. 103(10):3704–3709.
Springer MS, Meredith RW, Gatesy J, Emerling CA, Park J, Rabosky DL, Stadler T, Steiner C, Ryder OA, Janecka JE, et al. 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS One* 7(11):e49521.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28(6):1102–1104.
Sturm RA. 2009. Molecular genetics of human pigmentation diversity. *Hum Mol Genet*. 18(R1):R9–R17.
Szpiech ZA, Hernandez RD. 2014. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol*. 31(10):2824–2827.
Telfer PT, Souquière S, Clifford SL, Abernethy KA, Bruford MW, Disotell TR, Sterner KN, Roques P, Marx PA, Wickings EJ, et al. 2003.

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022

Molecular evidence for deep phylogenetic divergence in *Mandrillus sphinx*. *Mol Ecol*. 12(7):2019–2024.
Tosi AJ. 2017. Phylogenetic tests of a *Cercopithecus* monkey hybrid reveal X-chromosomal polyphyly of *C. cephus* and emerging patterns in the *cephus* species group radiation. *Afr Zool*. 52(3):177–181.
Tosi AJ, Detwiler KM, Clifford SL. 2006. X-chromosomal synapomorphies provide a non-invasive test for introgression among *Cercopithecus* monkeys. *Conserv Genet*. 7(5):803–805.
Tosi AJ, Detwiler KM, Disotell TR. 2005. X-chromosomal window into the evolutionary history of the guenons (Primates: *Cercopithecini*). *Mol Phylogenet Evol*. 36(1):58–66.
Tosi AJ, Melnick DJ, Disotell TR. 2004. Sex chromosome phylogenetics indicate a single transition to terrestriality in the guenons (tribe *Cercopithecini*). *J Hum Evol*. 46(2):223–237.
Tripathy V, Reddy BM. 2007. Present status of understanding on the *G6PD* deficiency and natural selection. *J Postgrad Med*. 53(3):193–202.
Tsiperson V, Gruber RC, Goldberg MF, Jordan A, Weinger JG, Macian F, Shafit-Zagardo B. 2013. Suppression of inflammatory responses during myelin oligodendrocyte glycoprotein-induced experimental autoimmune encephalomyelitis is regulated by *AKT3* signaling. *J Immunol*. 190(4):1528–1539.
Tung J, Barreiro LB. 2017. The contribution of admixture to primate evolution. *Curr Opin Genet Dev*. 47:61–68.
Tung J, Charpentier MJ, Garfield DA, Altmann J, Alberts SC. 2008. Genetic evidence reveals temporal change in hybridization patterns in a wild baboon population. *Mol Ecol*. 17(8):1998–2011.
van der Valk T, Gonda CM, Silegowa H, Almanza S, Sifuentes-Romero I, Hart TB, Hart JA, Detwiler KM, Guschanski K. 2020. The genome of the endangered dryas monkey provides new insights into the evolutionary history of the vervets. *Mol Biol Evol*. 37(1):183–194.
Vanhoutteghem A, Maciejewski-Duval A, Bouche C, Delhomme B, Herve F, Daubigney F, Soubigou G, Araki M, Araki K, Yamamura K, et al. 2009. Basonuclin 2 has a function in the multiplication of embryonic craniofacial mesenchymal cells and is orthologous to disco proteins. *Proc Natl Acad Sci U S A*. 106(34):14432–14437.
Vernot B, Akey JM. 2014. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* 343(6174):1017–1021.
Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, Dannemann M, Grote S, McCoy RC, Norton H, et al. 2016. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* 352(6282):235–239.
Wall JD, Schlebusch SA, Alberts SC, Cox LA, Snyder-Mackler N, Nevonen KA, Carbone L, Tung J. 2016. Genomewide ancestry and divergence patterns from low-coverage sequencing data reveal a complex history of admixture in wild baboons. *Mol Ecol*. 25(14):3469–3483.
Wang GD, Zhang BL, Zhou WW, Li YX, Jin JQ, Shao Y, Yang HC, Liu YH, Yan F, Chen HM, et al. 2018. Selection and environmental adaptation along a path to speciation in the Tibetan frog *Nanorana parkeri*. *Proc Natl Acad Sci U S A*. 115(22):E5056–E5065.
Wang H, Yin H, Jiao C, Fang X, Wang G, Li G, Ni F, Li P, Su P, Ge W, et al. 2020. Sympatric speciation of wild emmer wheat driven by ecology and chromosomal rearrangements. *Proc Natl Acad Sci U S A*. 117(11):5955–5963.
WHO. 2015. World malaria report 2015. *World Malar Rep*. 30:189–206.
Worley KC, Warren WC, Rogers J, Locke D, Muzny DM, Mardis ER, Weinstock GM, Tardif SD, Aagaard KM, Archidiacono N, et al. 2014. The common marmoset genome provides insight into primate biology and evolution. *Nature Genetics* 46:850–857.
Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 88(1):76–82.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–1591.
Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329:75–78.
Zhang M-L, Li M-L, Ayoola AO, Murphy RW, Wu D-D, Shao Y. 2019. Conserved sequences identify the closest living relatives of primates. *Zool Res*. 40(6):532–540.
Zhou X, Guang X, Sun D, Xu S, Li M, Seim I, Jie W, Yang L, Zhu Q, Xu J, et al. 2018. Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nat Commun*. 9(1):1276.
Zinner D, Fickenscher GH, Roos C. 2013. Family Cercopithecidae (Old World monkeys). In: Mittermeier RA, Rylands AB, Wilson DE, editors. Handbook of the mammals of the world. Vol. 3. Primates. Barcelona (Spain): Lynx Edicions. p. 550–627.

Downloaded from https://academic.oup.com/mbe/article/38/3/876/5912538 by guest on 26 January 2022